3D Passive-Vision-Aided Pedestrian Dead Reckoning for Indoor Positioning

Jingjing YAN, *Student Member, IEEE,* Gengen HE, *Assistant Professor, University of Nottingham Ningbo China,* Anahid BASIRI, *Lecturer, University College London,* and Craig HANCOCK, *Associate Professor, University of Nottingham Ningbo China*

***Abstract*—The vision-aided Pedestrian Dead Reckoning (PDR) systems have become increasingly popular, thanks to the ubiquitous mobile phone embedded with several sensors. This is particularly important for indoor use, where other indoor positioning technologies require additional installation or body-attachment of specific sensors. This paper proposes and develops a novel 3D Passive Vision-aided PDR system that uses multiple surveillance cameras and smartphone-based PDR. The proposed system can continuously track users' movement on different floors by integrating results of inertial navigation and Faster R-CNN-based real-time pedestrian detection, while utilizing existing camera locations and embedded barometers to provide floor/height information to identify user positions in 3D space. This novel system provides a relatively low-cost and user-friendly solution, which requires no modifications to currently available mobile devices and also the existing indoor infrastructures available at many public buildings for the purpose of 3D indoor positioning. This paper shows the case of testing the prototype in a four-floor building, where it can provide the horizontal accuracy of 0.16m and the vertical accuracy of 0.5m. This level of accuracy is even better than required accuracy targeted by several emergency services, including the Federal Communications Commission (FCC). This system is developed for both Android and iOS-running devices.**

***Index Terms*—altimetry, identification of persons, image processing, indoor environments, inertial navigation, position measurement, sensor fusion**

I. Introduction

DUE to unavailability of Global navigation satellite Systems (GNSS), e.g. the Global Positioning System (GPS) for indoor use, there have been a significant amount of researches to design and develop an alternative indoor positioning technology. They have resulted in several solutions, which can be divided into two main categories: infrastructure-based, and infrastructure-free [1]. Infrastructure based methods require costly and labor intensive pre-installations or regular management of related infrastructures. Meanwhile, the infrastructure-free methods overcome these limitations and are more promising, flexible, operational and marketable in the future [2]. In addition, the advancement of manufacturing common sensors used for infrastructure-free methods has also led to products with lower price, less energy consumption, smaller size and higher general precision [3-8]. Common examples are Inertial Measurement Units (IMU) of Micro-Electro-Mechanical Systems (MEMS) and cameras of Charged Couple Device (CCD). These advantages become more evident with the ubiquity of IMU sensors in smartphones and surveillance cameras in public building areas, leading to a wider range of applications in indoor scenarios in daily life [2]. However, none has yet provided a standalone solution that can provide continuous positioning at low or zero cost, and a multi-sensor system is considered to be a better option [6]. This paper, for the first time, uses the integration of Faster R-CNN based pedestrian detection by surveillance video, smartphone-based PDR, and barometer-based height/floor estimation to provide 3D positioning.

Using IMU sensors, PDR systems can provide the relative locations of users (not the absolute position though), the orientation, and the velocity of their movement. This can be potentially considered as a good solution for indoor use [4, 9-14]. PDR systems can be categorized into several groups with respect to where they are deployed and so what constraints can be applied. They include the foot-mounted [15-20], waist-mounted [21] and hand-held systems [22-26]. This study proposes and implements a novel PDR system for handheld smartphones, to make most of their wide use and ubiquity [27, 28], and also the miniaturized and low-cost sensors that are embedded in the phone [1]. However, the accumulating temporal drift is still the major challenge for many applications [20, 29, 30]. It will accumulate with time and the positioning errors may exceed 100m in 1 minute [9]. This leads to errors in long-term PDR-alone positioning, and thus external positioning information is required for position calibration and absolute localization [4, 11, 31-33].

Manuscript received Nov, 9th, 2018. The author acknowledges the financial support from the International Doctoral Innovation Centre, Ningbo Education Bureau, Ningbo Science and Technology Bureau, and the University of Nottingham. This work was also supported by the UK Engineering and Physical Sciences Re-search Council [EP/L015463/1].

Jingjing. Yan is with International Doctoral Innovation Centre, University of Nottingham, Ningbo, 315000, China (e-mail: Jingjing.YAN@nottingham.edu.cn).

Gengen He is with Department of Geographical Science, University of Nottingham, Ningbo, 315000, China (e-mail: Gengen.He@nottingham.edu.cn).

Anahid Basiri is with Centre for Advanced Analysis, University College London, WC1E 6BT, London (e-mail: a.basiri@ucl.ac.uk).

Craig Hancock is with the Depart of Civil Engineering, University of Nottingham, Ningbo, 315000, China, (e-mail: Craig.Hancock@nottingham.edu.cn).

One solution for this can be multi-sensor fusion, i.e. the integration of additional sensors [31, 33]. In recent researches, approximately two-thirds of multi-sensor systems are inertial systems calibrated by external systems, and their common calibration choices are Received Signal Strength (RSS), Time-of-Flight (ToF), and map matching [34]. This study chooses Optical Positioning System (OPS), which is under-represented in previous studies [34]. However, it has a huge potential due to its relatively higher accuracy with increasing availability in the form of surveillance videos [1, 35]. The reason of not having a significant amount of work on OPS could be explained by its performance vulnerability with the possibility of obstruction of the Line-of-Sight (LoS) signals (between the camera and the targets), which is common inside the buildings, and generally indoors [36, 37]. This may result in failure of the OPS to provide a reliable and continuous positioning solution. While this has remained as one of the major challenges of indoor tracking and navigation [38], despite several studies trying to predict the pedestrians' positions by using a wide range of algorithm including Kalman filters [39-41] and linear regression [42, 43], this paper uses OPS alongside with PDR to overcome the LoS challenge. This is mainly due to the fact that predicting the pedestrian's location can be unreliable due to the unpredictability of human movements. The introduction of OPS in the hybrid positioning can also enrich information gathering from visual data by object detection [8, 35].

Some recent studies [11, 22-24] have proved that the integration of these two sensing systems, also known as Vision-aided Inertial System (VINS), can keep the advantages of both positioning systems while providing 2D localization service with higher accuracy, continuity, accessibility, and reliability [13, 33]. However, the previous passive VINSs (PVINSs) only focus on providing the 2D positions in a fixed scene with a single camera using self-trained pedestrian detectors [11, 22-24]. To overcome these limitations, this study contributes the first multi-camera 3D PVINS, with the ability of handling multi-scene shifting. Its algorithm for pedestrian detection, i.e. Faster R-CNN, can be directly implemented without training by utilizing online resources while achieving real-time detection with high detection accuracy. Meanwhile, it contributes a novel algorithm for automatic scene-shifting integrated with PDR's automatic turning recognition. Second, this study provides a simple but effective novel algorithm to integrate PDR and visual tracking systems. Its performance has achieved at least 20% accuracy improvement of synthesized results in an environment with over 65% entirely invisible areas, when the previous 2D PVINSs are applied in environments with less than 50% partial occlusions [22-24]. Third, it contributes a novel algorithm to detect different floors and even their transition areas as the 3D information by using a smartphone-embedded barometer. Fourth, it is the first study which presents the acquired results in the automatic switching floor plans with absolute world coordinates and they can be directly used in outdoor systems. Finally, all previous studies are only tested on Android systems, while this study is the first one using both smartphones with iOS and Android operating systems.

This paper contributes a novel design of a 3D PVINS with relatively higher accuracy. It tracks 2D user movements of each floor by using multi-cameras and smartphone-based PDR while identifying the current user floor and height by using a smartphone-based barometer. The prototype will be tested in a four-floor building by using two types of smartphones, running the commonly deployed iOS and Android operating systems. The paper is organized as follows. Section II compares the details of methods used in this study and other studies. Section III describes the components of the system used in this study. Section IV introduces the experimental design and Section V presents the results with the comparisons to other methods, and Section VI presents the conclusion.

II. Related Works

Based on the way of system deployment, the VINSs can be divided into two classes: the active VINSs (AVINSs) and the PVINSs. The AVINSs have been used extensively. They can provide 3D location information and orientation estimation for motion tracking [37]. Potential applications include robotic navigation, Simultaneous Localization and Mapping (SLAM), and unmanned vehicle systems. Their common setup is to attach a single camera and IMU sensor together on a fixed platform [12, 31, 44, 45]. In order to integrate the Inertial System (INS) and video, common methods include Particle Filter (PF) [2, 46], Kalman Filters (KF) [13, 47] and its extensions such as Extended Kalman Filters (EKF) [7, 36] and Unscented Kalman Filters (UKF) [31].

Some of AVINSs have utilized the embedded cameras and IMU sensors in smartphones for indoor localization [48-50]. However, this approach is not fully practical, particularly for commercial applications, as video recording by the embedded camera is energy consuming and cannot support long durations for indoor localization. Therefore, the authors previously suggested deploying surveillance cameras for pedestrian detection while using inertial sensors in smartphones [51, 52].

This method is regarded as PVINS. Other than AVINS, the sensors in this system are distributed on different platforms and further data transformation is needed before sensing integration. Some of the recent studies utilize this idea to provide 2D locations [22-24]. These studies integrate the visual results from a single surveillance camera and PDR results from the embedded IMU sensors in the smartphone to continuously track 2D user movements in either indoor or outdoor environments. The studies conducted by Missouri University [22, 23] apply similarity matrices to combine the visual and PDR trajectories by checking whether the distance between these two trajectories is within a certain threshold in each sliding window. For visual tracking, it tracks the user in the visible areas by self-trained SVM-based detector. The acquired results from the filming view are warped to a top-down view by using four corresponding pairs. This requires the whole filming scene to be fixed, and be covered inside the visible area of the camera. For PDR positioning, it is based on speed vector with fixed step length and moving direction, by using accelerometer, gyroscope, and magnetometer. Both positioning results from these two sub-systems are required to be transferred to relative world coordinate for trajectory matching. This system can

achieve bi-directional calibration for both PDR and visual results. However, as the similarity matrix in a corresponding sliding window needs to be updated during the matching process, it may lead to some difficulties in computation. It may also require some more computation when shifting to a second camera as the warping matrix needs to be re-calculated. In addition, it may have detection-lag errors to determine whether pedestrian detection is still working by checking the frames in a certain duration. Moreover, the final positioning results are still in relative coordinates and they do not provide a solution to link them to real geographical coordinates. In this paper, the integration of visual positioning and PDR is based on the PDR's heading calibration from visual orientation instead of using their positions. Therefore, the system operation is simpler as it does not need to calculate these matrices and visual tracking process can freely shift from one camera to another as a multi-camera system. Moreover, as this study applied deep-learning methods for pedestrian detection, it does not need a self-training process as the detectors are already available resources. Meanwhile, it does not need self-updating of scales, as the algorithm can automatically update the detector's size and it can save some manual work. It also achieves nearly real-time detection and it can respond immediately with zero detection. Therefore, it can reduce delay-detection errors of the system.

Another study conducted by Shanghai Technology University [24] combines the two systems by matching the gait features from both visual and PDR system. For visual tracking, their system also installs the camera to view the whole scene. In no-occlusion areas, it uses foreground segmentation for pedestrian detection. The detected user feet position in each frame is on the extension cord of two points: the top point of the foreground mask and the gravity center of the bounding box (BB). The occlusion area in that study is defined as the condition that the pedestrian is only partially detected. In these areas, the feet point is regarded as the mid-point of the bottom boundary detected by Convolutional Neural Network (CNN). For visual gait feature extraction, it is achieved by finding the repeating pattern of a higher proportion of the lower body in BBs. After combining step state, step frequency and heading, the gait features from two systems with the largest matching rate will be integrated for 2D positioning. This method can improve feet position accuracies in no-occlusion areas. However, it increases the responding time of system as it needs more processing steps and the foreground segmentation method cannot process pedestrian detection as quickly as deep learning does. Moreover, this algorithm cannot be applied in areas when people are too close to the camera. It will limit the system accuracy as the feet points cannot be treated as the bottom mid-points as their feet are invisible in the scene. In this paper, the filming areas have no occlusions, and thus the system does not need to separately treat the calculation of feet positions. Moreover, it removes those BBs when no entire human bodies can be viewed in the frames. Comparing the matching algorithms, the method in [24] needs gait feature extraction before integrating visual tracking and PDR data together, leading to an increase of the computation complexity for the application. In this paper, it only needs the similarity checking of time stamps from two sub-systems as it is a continuous process and it is simpler to achieve. Meanwhile, the system proposed by [24] still does not provide a solution to transfer the positioning results to the absolute geographical coordinates, i.e. the global mapping system. This paper has solved that problem and it can provide opportunities for further application of seamless indoor-outdoor transition. This paper also compares the performances between two common types of smartphone models. Other than only using Android-running smartphones in the previous studies, it has improved the system robustness for different kinds of smartphones.

This study first improves the design of the previous system in the aspect of 2D PVINS. The video data are used to be derived from a single camera within a fixed scene [51-53]. However, in this study, they are acquired from multiple cameras with scene shifting to enlarge the visible areas for continuous tracking of user movement. Moreover, the sensor fusion method previously used is based on position replacement by time synchronization, which is not that close to reality. This study employs an alternative approach called heading calibration based on the comparison of the accuracy of these two methods conducted in [52]. It also adds the step length calibration to improve the performance of the PDR system. With these improvements, the 2D accuracy (0.16m) of this system is significantly higher than the best performances provided by the Commercial Mobile Radio Service (CRMS) reported in FCC (5~10 m) [54]. Moreover, the previous studies [11, 22-24, 51-53] only provide 2D user locations, while to enable a continuous positioning service, particularly for the time the user is walking up or downstairs, a 3D (or the recognition of the floors) are required [55-57].

This paper introduces embedded barometer from smartphone and provides the height and floor estimation by contributing a novel floor detection algorithm with the integration of pre-stored camera locations. It achieves a vertical accuracy of 0.5m with 98% accurate floor detection, which is significantly better than the requirement of FCC (3m) [54]. Some of the previous studies use IMU sensors to provide heights. However, it will still raise the problem of increasing bias in vertical direction [57-59], due to the introduction of nonlinearity caused by accelerometer rotation during measurements. The errors will grow quadratically with time and they cannot be handled efficiently by standard EKF [29, 58]. Thus, the fusion of other sensor data is necessary to stabilize the height tracking by fixed beacons or data training [29, 30, 60]. The former one will have additional cost for installation [61], and the latter one requires a high-cost data training process [57], leading to relatively high energy consumption [57, 62].

Using a barometer may be a good alternative solution [61]. First, it has been widely used at outdoors for altitude measurements [63, 64], as it is low in energy cost [57, 64-66] and requires no additional installations. Second, more smartphones are now embedding pressure sensors, such as Galaxy Nexus4, Samsung S4, iPhone6, Xiaomi Mi2, and their more recent versions [56, 57, 61, 64, 66, 67]. Together with corresponding software for data fusion, the portable sensor-assisted methods have drawn more attention in the field of

height tracking [30, 56, 57, 64].

MEMS barometer can be integrated with IMU sensors, which is known as baro-IMU for indoor navigation system [30, 55, 58, 68, 69]. It can improve the accuracy of providing height information than using only MEMS accelerometers [59, 66]. For example, one previous study has loosely coupled these two types of data with self-designed hardware under experimental conditions. Its height estimation has achieved an RMSE in a range between 0.05m and 0.68m with simple motions [30]. A later study [68] applies this approach with smartphone sensors to help guide the blind in subway stations and commercial centers with longer distance, achieving decimeter-level accuracy on height estimation. However, many studies more concentrate on improving the 2D positioning accuracy by enhanced PDR algorithm, rather than focusing on the vertical height error. They just collect the pressure data of each floor as fingerprints and treat the between-floor height as constant, with a pre-calibrated pressure sensor by GNSS signals [69, 70]. The typical vertical error is approximately 2m [69], and the detection accuracy is still unknown as they do not provide any results about whether the floor detection can be performed accurately and in time. This may be explained by that the requirement for floor detection by barometer is not very high in the real-world applications, as the height difference between floors is relatively significant. This paper introduces the transition levels between floors during measurements, which is usually neglected in previous studies [30, 55, 58, 68, 69]. Therefore, the accuracy of height estimation is becoming more important as more detailed height changes are needed. Some studies set up a referential device to improve the height estimation. They have achieved better mean accuracy at about 0.15m [71]. This study also adheres to the idea of providing height information for indoor tracking. However, it only uses a single device but different data collection tools to set up referential measurements. In addition, as the barometer can only help improve the performance in the 3rd dimension [61], it still needs an external positioning system for calibration on the 2D aspect, which corresponds to the PVINS in this study.

To sum up, this study contributes a novel design of a 3D indoor tracking system with the integration of the passive multi-scene OPS, active PDR and altimetry estimation, supported by auto-shifting georeferenced maps. It is the first time to use only these three sub-systems for 3D localization simultaneously and collaboratively.

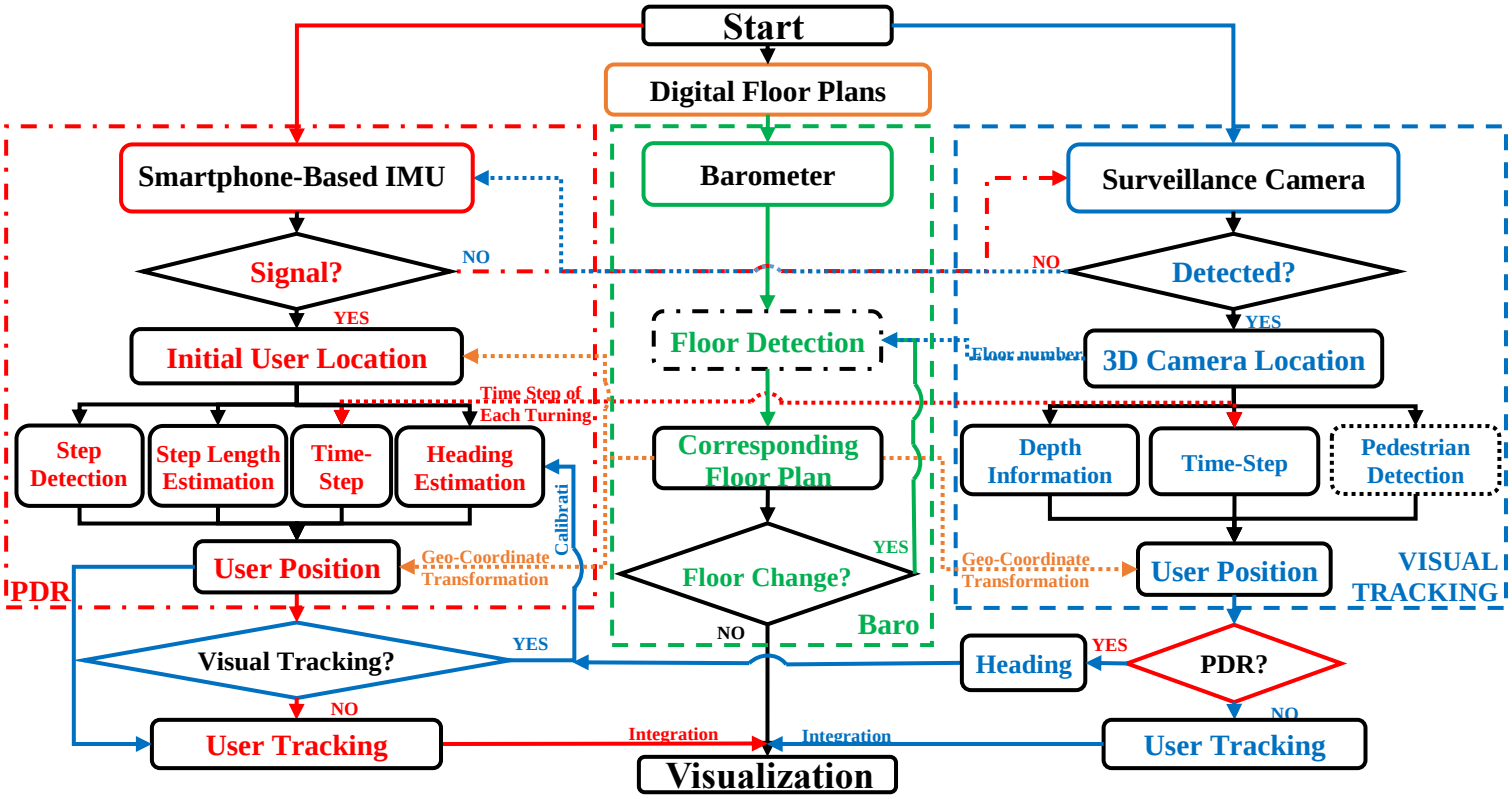


Fig. 1. The architecture of the proposed system (PDR, Visual, Barometer, and digital floor plans are represented in red, blue, green and orange, respectively). The pedestrian detection and floor detection in dash-boxes will be explained later in Section III B and III D, the Geo-Coordinate Transformation is in Section III C).

III. System Design

The structure of our system is shown in Fig.1. During the operation, the smartphone-based PDR keeps actively tracking the user movement, while the OPS only functions in the LoS areas, shifting from one scene to another. This is an update to previous work [51-53], as it can handle multi-scene instead of a fixed scene. During the movement, the smartphones are held horizontally and pointing forward. The accelerations and the angular velocities are collected simultaneously. The former is used for step detection and step length estimation while the latter is applied for heading estimation. The integration of these data can help calculate relative 2D PDR positions. Meanwhile, the video recording is triggered since the user starts moving. Once entering the LoS area of each camera and a significant change is detected from the estimated PDR headings, the 2D

visual positions will be calculated based on BBs' positions by pedestrian detection and the estimated depth information in corresponding frames. Meanwhile, the 3D information of the corresponding functioning camera will also be reported to the main system, which will help to calibrate the floor detection. The 2D visual headings are determined by visual positions in every two consecutive frames [52]. They will later be used for 2D PDR heading calibration based on similar time stamps. For data fusion, this study replaces previous time-synchronization-based position replacement in [51] by using synthesized results from calibrated headings and PDR step lengths. This is because heading calibration responses better to the real-world scenarios based on the conclusions in [52], thus it can provide better synthesized position estimation.

Before 2D calibration, the 2D results from PDR and visual tracking should both be transformed into real geographical coordinates, i.e. geo-coordinate transformation. It is beneficial for further development of seamless indoor-outdoor positioning [51, 52]. To achieve that, the corresponding floor plans will help to provide absolute positioning information. These maps are pre-stored in the system and will be integrated into the 2D PVINS results by automatic selection based on the results of floor detection. The system in 2D PVINS aspect is providing a calibrated 2D path in an absolute coordinate system at each epoch, i.e. the corresponding time stamps of each step. This 2D path will later be integrated with the estimated height and floor information by finding the similar time stamps.

For 3D information, the system uses the smartphone-based barometer to continuously identify the current floor of the user during movements. Before the start of operation, the barometer needs a self-calibration. This is done by comparing and adjusting the two readings acquired from two smartphone apps installed on the very same smartphone. One is chosen as the 'standard' pressure, and the other is calibrated measurement for the same reading. The calibrated measurements are then processed for height estimation and floor detection. This process will be discussed in more details in sub-section III D. With the pre-stored 3D locations of cameras in the system, the calibrated results for the floor detection can also be improved.

Having distinguished the floors, the results will be integrated with 2D PVINS with a minimum difference of time stamps. The final 3D path will be presented in a 2D form on each floor with the corresponding georeferenced floor plan for visualization.

A. 2D Smartphone-Based PDR

The proposed inertial positioning proceeds as follows: (1) step detection, (2) step length estimation, (3) heading estimation, and (4) position estimation. This paper improves the previous works [51, 52] by calibrating estimated step lengths and adds automatic turning detection in the heading estimation. This can improve the 2D PDR positioning accuracy and assist data fusion with visual tracking.

1) Step Detection

The measurements from the accelerometers are first filtered using a low-pass filter with frequency condition as a function of the accelerometer's sampling rate [5]. Then, the raw motion accelerations with respect to time taken in three axes as $a_x(t), a_y(t)$, and $a_y(t)$, needs to be synthesized. This is due to the distribution of vertical signals, which mainly contributes to the step peaks and can appear in all axes depending on the current device's altitude and orientation [52, 72]. While may not be always true, but the projection to the horizontal axis can be done. In addition, the training of evacuation may include such recommendation to the users. Having assumed the horizontal grip, the step detection is only related to the relative synthesized motion accelerations in the vertical direction $a^*(t)$ and its magnitude can be calculated as in (1):

$$|a^*(t)| = \sqrt{(a_x(t))^2 + (a_y(t))^2 + (a_z(t))^2} - g \qquad (1)$$

where g is the earth's gravity, requiring to be removed from the vertical motion component. The synthetic motion's magnitude $|a^*(t)|$is then needed to be processed by applying a pre-settled threshold to identify different features of a gait cycle in each sliding window as one acceleration, two static and one deceleration phase. The length of the window is determined by the frequency of the accelerations. After that, a zero-crossing approach is then applied to detect different cycles i [51, 73].

2) Step Length Estimation

Step length estimation is based on Weinberg's algorithm as demonstrated in (2), which uses a non-linear model with maximum ($|a^*{}_{max}(i)|$) and minimum ($|a^*{}_{min}(i)|$) of synthetic accelerations' magnitude of each step event i [51, 74].

$$SL_i = \sqrt[4]{|a^*{}_{max}(i)| - |a^*{}_{min}(i)|} * k \ (i = 1,2, \dots, n) \qquad (2)$$

where SL_i is the step length of the i^{th} step and k is an empirical value of penalty for estimation [51], which can be determined by the ratio between processed results of accelerations and assumed walking step length in 1.22m concluded from previous studies in [75-77]. Then SL_i will be calibrated by a ratio η which is determined by the sum of estimated step length and the length of reference path D_{Real} in 2D, which is not processed in previous work.

$$SL'_i = \eta * SL_i \,, \eta = \frac{\sum_{i=1}^{n} SL_i}{D_{Real}} \ (i = 1,2, \dots, n) \qquad (3)$$

3) Heading Estimation

Each step's orientation is relative by its corresponding angular velocity changes in the body frame, which can be measured by the embedded three-axis gyroscope in a smartphone as ω_x^t, ω_y^t and ω_z^t [9]. The changes of heading in body frame from the current stage to the next stage within certain duration Δt can be described as:

$$\Omega = \begin{pmatrix} 0 & -\omega_z^t \Delta t & \omega_y^t \Delta t \\ \omega_z^t \Delta t & 0 & -\omega_x^t \Delta t \\ -\omega_y^t \Delta t & \omega_x^t \Delta t & 0 \end{pmatrix} \qquad (4)$$

The next step is to transfer that change from body frame to the global frame by using a 3×3 rotation matrix as:

$$R(t + \Delta t) = R(t) * \exp(\Omega) \ (t > 0) \qquad (5)$$

where $R(t)$ is the rotation matrix of the current stage and the $R(t + \Delta t)$ for the next stage. When in the initial stage, the rotation matrix $R(0)$ can be described by rotations happened in three axes as $R_x(0), R_y(0), R_z(0)$. The transformation process is described in (6)-(9):

$$R_x(0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi(0)) & -\sin(\phi(0)) \\ 0 & \sin(\phi(0)) & \cos(\phi(0)) \end{pmatrix} \qquad (6)$$

$$R_y(0) = \begin{pmatrix} cos(\theta(0)) & 0 & sin(\theta(0)) \\ 0 & 1 & 0 \\ -sin(\theta(0)) & 0 & cos(\theta(0)) \end{pmatrix} \quad (7)$$

$$R_z(0) = \begin{pmatrix} cos(\psi(0)) & -sin(\psi(0)) & 0 \\ sin(\psi(0)) & cos(\psi(0)) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (8)$$

$$R(0) = R_z(0)R_y(0)R_x(0) \quad (9)$$

where $R_x(0), R_y(0)$, and $R_z(0)$ are sub rotation matrix consist of roll $\phi(0)$, pitch $\theta(0)$ and yaw $\psi(0)$ directions of body frame respectively. The overall rotation matrix $R(0)$ is determined by the integration of these three components. The initial states of roll $\phi(0)$ and pitch $\theta(0)$ angles are determined by average changes of initial accelerations in corresponding directions and the initial yaw $\psi(0)$ will be zero [52]. The next step is to find corresponding Euler angles from calculated rotation matrices $R(t)$ from angular velocity changes. In this study, as the smartphone is held in a relatively stable condition by user's hand, pointing to the walking direction, the heading i.e. $\psi(i)$ of each step is only the results of changes in yaw direction [5, 52] and can be calculated as in (10) based on the previous detected step events i:

$$\psi(i) = \arctan2\left(R_{2,1}(i), R_{1,1}(i)\right) \ (i = 1,2,\dots,n) \quad (10)$$

This is because the rotation matrix $R(t)$ can be rewritten as:

$$R(t) = \begin{bmatrix} R_{1,1}(t) & R_{1,2}(t) & R_{1,3}(t) \\ R_{2,1}(t) & R_{2,2}(t) & R_{2,3}(t) \\ R_{3,1}(t) & R_{3,2}(t) & R_{3,3}(t) \end{bmatrix} \quad (11)$$

and we have $\begin{cases} R_{2,1}(t) = \cos(\theta(t))\sin(\psi(t)) \\ R_{1,1}(t) = \cos(\theta(t))\cos(\psi(t)) \end{cases}$ (12)

According to these equations, it can be found out that:

$$\tan(\psi(t)) = \frac{R_{2,1}(t)}{R_{1,1}(t)} \quad (13)$$

As the path in this study is more complicated than in previous works [51, 52], the acquired headings $\psi(i)$ needs to be processed for automatic turning detection by finding the sudden changes of average values with a certain threshold applied (Fig.2).

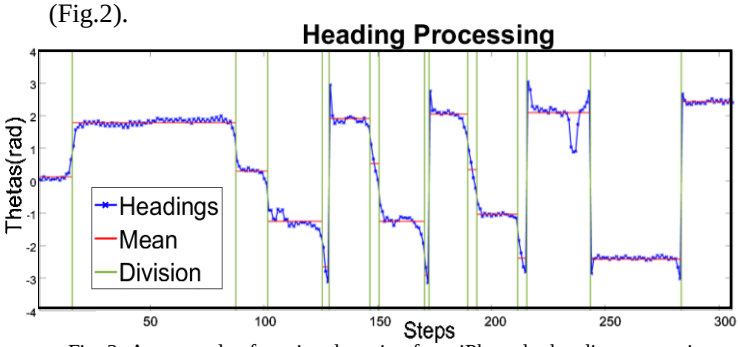


Fig. 2. An example of turning detection from iPhone by heading processing.

It can later be used for matching with visual tracking for 2D position calibration. The previous study has tried to extract features from both magnetometer and gyroscopes for heading direction classification by applying Principal Component Analysis (PCA) algorithm [70]. However, this will increase the computation complexity, while introducing some unexpected detection errors. This will also reduce the variety of headings to eight directions. The method used in this study tries to simplify the computation process by using only gyroscope, smooth down these unexpected changes in headings by averaging while providing more options for heading directions. In this study, about 16 corners are detected and their average delay of detection is approximately one step.

4) Position Estimation and Error Calculation

The user position P_i is calculated by the combination of corresponded estimated step length SL_i with estimated heading $\psi(i)$ and the location of the previous step:

$$P_i = \begin{bmatrix} P_{E_i} \\ P_{N_i} \end{bmatrix} = \begin{bmatrix} P_{E_{i-1}} + SL_i * \sin(\psi(i)) \\ P_{N_{i-1}} + SL_i * \cos(\psi(i)) \end{bmatrix} \quad (14)$$

where P_{E_i} and P_{N_i} represent the eastern and northern position components separately [5, 52]. Before the calculation of position error, the estimated positions need to be transformed into a real geographic system as the reference positions are measured in this way [52]. The positioning error $E_i$is defined as the distance between the estimated position P_i and reference position R_{f_i}, and is calculated in (15). Then the Mean Average Error (MAE) is calculated as the mean of all E_i in (16).

$$E_i = \left\| P_i - R_{f_i} \right\| \quad (15)$$

$$\text{MAE} = \frac{1}{n}\sum_i^n E_i \ (i = 1,2,\dots n) \quad (16)$$

B. 2D Visual Tracking

1) Pedestrian Detection by Faster R-CNN

Common methods for passive pedestrian detection are feature-based segmentation [78-82]. They usually require manual selections for the best features, leading to constraints in the applications and the test environments. This is mainly due to their parameters of algorithm needs to be modified regularly, which is affected by the ambient. This study uses a deep-learning based method for pedestrian detection, called Faster R-CNN [51]. It requires a minimum of manual inputs as almost the whole process is automatic, providing a solution with relatively high flexibility and ubiquity [83-86], in comparison with the feature-based methods. It is based on 3-layer Regional Proposal Network (RPN) and 5-layer Region-Based CNNs (R-CNNs), and is one of the state-of-art methods for deep learning with higher accuracy and real-time processing [51-53, 86]. The RPN is used for recognizing the potential object areas (ROIs). The ROIs are processed by Pooling for BB prediction and the results are passed to Full-Connected layers for later Softmax classification to differentiate all classes. This study simplifies the original 20 classes into two: 'human' and 'non-human'. Meanwhile, the BB regression is used to improve the detection accuracy (Fig.3). Some of the later studies have attempted to increase the robustness of Faster R-CNN by improving the performance of detecting partial human bodies [87], however, this study mainly focuses on the detection of entire human bodies and thus still uses Faster R-CNN.

In this study, the cameras are located on different floors and they are pointing nearly orthogonally to the corridors. As the resolution of the camera is too low for facial recognition, there is no risk of privacy infringement. Before the operation, the acquired video data need to be divided into frames for later processing as the Faster R-CNN algorithm only works for individual images. These frames will be uploaded to the system by streaming. Once the user is detected by the system, it will

first recognize the functioning camera with corresponding location information. After being processed by Faster R-CNN, the BBs are extracted from these frames and the corresponding frame numbers are also recorded for later time stamps acquisition. As the size of these BBs will be automatically adjusted to the size of device holders in the frames and the cameras are facing nearly orthogonally to the corridor, the gravity center of the filming user is therefore assumed to be at the center of BBs. The middle points of the BBs' bottom boundaries are then regarded as the lowest points of the users or potential user's mobility aid [51-53, 86]. Their coordinates can be determined by the horizontal coordinates (x_i^1, x_i^2) of the BBs in frames, which needs to be transferred to the real distance by the width of the BBs and the width of each frame, and the depth information D'_i which is derived from a pinhole model. These points can be constructed into the entire user path (Fig.4).

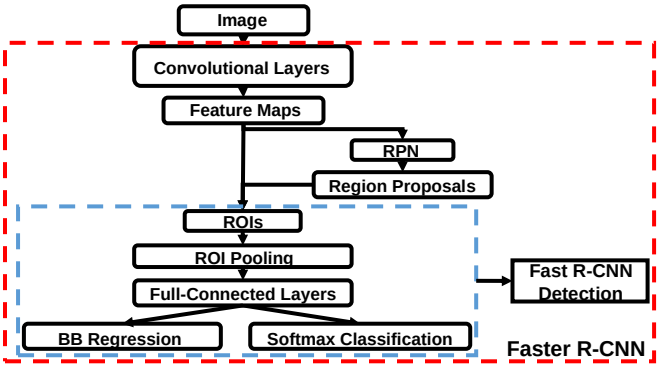


Fig. 3. The framework of Faster R-CNN (ROI: Region of interest).

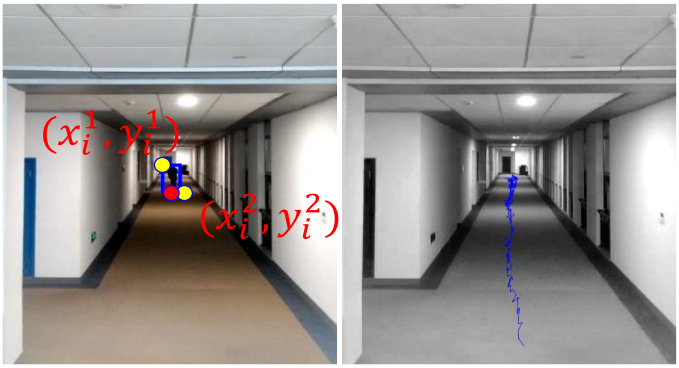


Fig. 4. An example of extracted BB from the frame (left) and entire user path (right), where (x_i^1, y_i^1) represents the upper left of BB, and (x_i^2, y_i^2) reprents the lower right of BB.

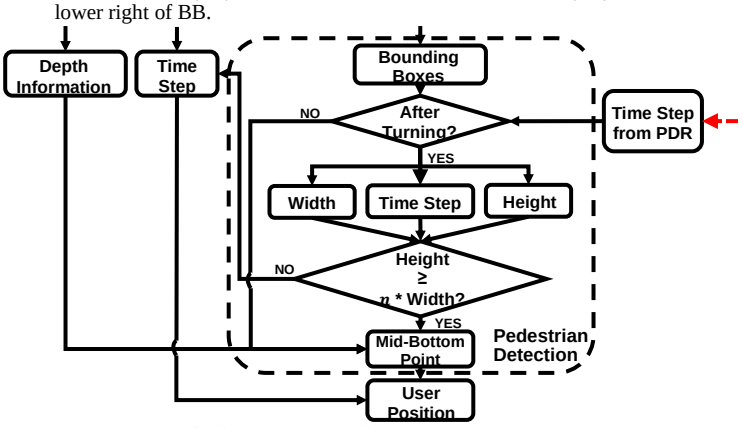


Fig. 5. The shift of visual tracking from one camera to another.

This process only functions with both a sudden change of PDR's headings (i.e. at the corners) and the average ratio n of height and width of extracted BBs ($n = 2.5$ in this study) (Fig.5). This is an update to the previous works of single-scene [51-53]. It allows the multi-scenes filmed by different cameras can be shifted from one to another based on PDR's time stamps. It also helps to remove incorrect measurements of pedestrian detection caused by long filming distance [51-53] and ensures the entire human body is included in each BB for the steps.

2) Person Localization

The coordinate of the user in each frame can be represented as $(X_i, D'_1 - D'_i)$, where X_i is related to the ratio between the width of the frame and the corridor:

$$\frac{\frac{(x_i^1 + x_i^2)}{2}}{X_i} = \alpha * \frac{W_F}{W_{Corr}}, \alpha = \frac{\sqrt{(x_i^2)^2 - (x_i^1)^2}}{W'_{Corr}} \quad (16)$$

where (x_i^1, x_i^2) is the horizontal coordinate of the upper left and lower right corner of each BB, W_F is the frame width, W'_{Corr} is the relative corridor width on the user position in the frame, $W_{Cthe\ orr}$ is the real width of the corridor, which will be provided later by integrating map information. The depth information $D'_i$is driven from the distance D_i between user and camera in i^{th} frame, which could be determined by a pinhole camera model [80] as in (17) with a pixel height h_i, focal pixel length f and real height H_p of human. f is determined by camera resolution and field of view (FOV) [51-53].

$$\frac{h_i}{f} = \frac{H_P}{D_i} \ (i = 1,2, \dots, n) \quad (17)$$

The real depth information can be then driven according to the filming mechanism of the camera (Fig.6) as:

$$|D'_i| = \sqrt{(D_i)^2 - (H_C - H_P)^2} \quad (18)$$

where H_C represents the height of the camera which is 3m, and H_P represents the height of the person, which is 1.6m in this paper. The first depth gathered from the calculation D'_1 can be calibrated by the real length of the corridor, which will be provided by the map information and the ratio γ will be applied to the calculation of $|D'_i|$:

$$|D'_i|^* = \gamma * |D'_i| \quad (19)$$

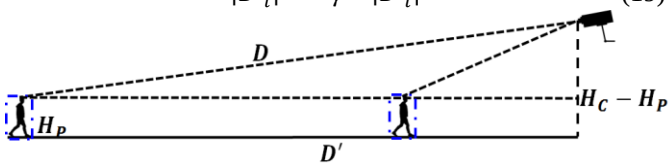


Fig. 6. The filming mechanism of a camera for depth information calculation.

The heading information is subsequently determined by detected user position points from every two consecutive frames as $(X_i, D'_1 - \gamma * D'_i)$ and $(X_{i+1}, D'_1 - \gamma * D'_{i+1})$.

C. 2D Floor Plan Integration and Heading Calibration

1) 2D Floor Plan Integration

Before calibration, both 2D results achieved from smartphone-based PDR and camera-based visual tracking need to be projected into the same coordinate system provided by map information, i.e. geo-coordinate transformation. The way to achieve that is by applying rotation M, scaling β, and translation δ:

$$\begin{bmatrix} x_R \\ y_R \end{bmatrix} = M \begin{bmatrix} x_r \\ y_r \end{bmatrix} * \beta + \delta \quad (20)$$

where (x_R, y_R) are the coordinates from the real global

geographical coordinate system while (x_r, y_r) are from the relative world coordinates, which can be (P_{E_i}, P_{N_i}) from PDR and $(X_i, D'_1 - \gamma * D'_i)$ from visual tracking systems. The rotation M, scaling β, and translation δ can be determined during the process of geo-referencing by finding three pairs of points on the floor map between relative and absolute positioning system.

This study digitizes the image floor plans to digital maps by geo-referencing. The absolute positions with some simplified semantic representations of indoor building information are then created. The digitized floor plans of different floors are georeferenced into WGS84 coordinate system with assigned floor level. With the assistance of floor detection during the movement, the corresponding floor plan will be automatically selected for visualization. The use of WGS84 would help to develop a seamless transition between the indoor and outdoor environments. This is particularly helpful as it is a widely used Spatial Reference System (SRS) for GPS and many other outdoor positioning technologies using the same SRS [51-53].

2) 2D PDR Heading Calibration Based Data Fusion

For calibration, this paper uses headings from visual positioning results to calibrate horizontal drifts of PDR-based positioning. In our previous work, a method called position-replacement is used, however, our research later compared it with heading calibration and found the latter preferable due to higher accuracy [52]. This is because the sampling frequencies of these two positioning systems are different, and the positioning results from visual tracking cannot be directly applied to visual tracking. This paper uses heading calibration, which replaces each step's heading $\psi(t(i))$ acquired from PDR by visual orientation decided by two consecutive frames with similar time steps. The time steps of PDR are deduced from the detected step events and the related time stamps from the accelerometer readings, while that of the videos are inferred from the frame number and filming frequency.

$$\psi(t(i)) = \lim_{(t(j)-t(i))\to 0} \arctan_2\left((X_{t(j)}, D'_1 - \gamma * D'_{t(j)}), (X_{t(j+1)}, D'_1 - \gamma * D'_{t(j+1)}\right) \quad (21)$$

where $t(i)$ is the time step from the i^{th} step event, $t(j)$ and $t(j+1)$ are the time steps from j^{th} and its following frames. The 2D user positions will be recalculated based on the integration of these calibrated headings and pre-calibrated step length SL'_i [52]. With this process, PDR and OPS are integrated together to provide a 2D path, with synthesized headings from PDR and OPS, and calibrated step length from PDR.

D. Floor Detection by Barometers

A barometer altimeter allows height estimation based on air pressure above the given reference level, which is by default sea level [30, 56, 63, 64]. The equation used in this study is in (22):

$$h = \frac{((\frac{P_0}{P})^{\frac{1}{5.257}} - 1) * T}{0.0065} \quad (22)$$

where P_0 is the standard atmospheric pressure, which is 101.325 kPa, P is the absolute pressure, and T is the temperature in K.

However, the pressure information acquired by single barometer can be very noisy. It can be easily affected by the ambient factors, such as temperature, humidity, time, and even opening and closing of windows or doors. On the other hand, the relative changes of pressure between floors can be regarded as a constant value [30, 61, 63, 64, 66]. One study tried to build up fingerprints for relative heights of each floor, but it is time-consuming and the accuracy is not satisfied as 1 to 2m [69]. In a later study [71] uses one reference device and one carrying device for exact floor identification and reaching an accuracy of 0.15m. This study contributes a novel algorithm to provide the 3D information by using only single barometer while the previous works have no such information [11, 22-24, 51-53].

This study does not need the installation of a reference device. Instead, it uses a single smartphone-based barometer but along with two barometer apps: 'Barometer' can provide pressure at the ground level and the current level and 'Barograph' keeps recording pressure reading during movement. Their readings will be compared initially for self-calibration and then the relative height ΔH_i is then calculated by the removing the effect from ground level:

$$\Delta H_i = h_i - h_g \quad (23)$$

The pressure data will be processed by finding the sudden changes in trend and average for floor change detection (Fig.7). Instead of using a certain value for each floor, which has been used in many studies before [57, 64, 71], it uses a referential height range $(\Delta H_{R_{j(Bottom)}}, \Delta H_{R_{j(Top)}})$ for each floor to deal with variations caused by multiple unknown factors.

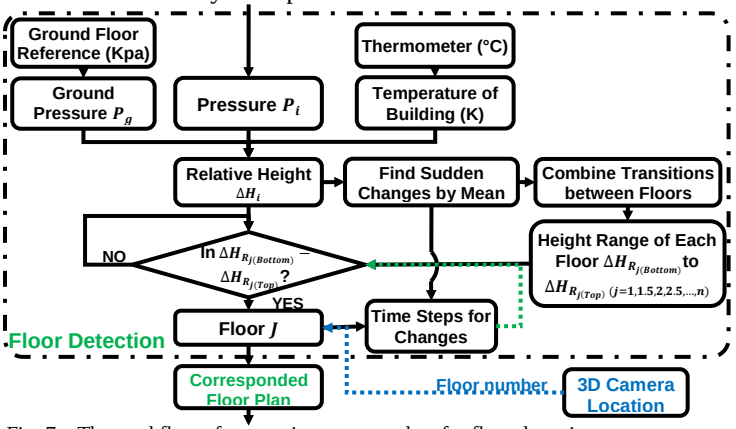


Fig. 7. The workflow of processing pressure data for floor detection.

The indoor temperature is supposed to be measured by an indoor thermometer, which needs to be pre-installed in the building. However, the indoor temperature in this experiment is controlled by an air condition system and it can be regarded as thermostatic, with the estimation of 20ºC (293.15K).

The algorithm also treats the transition areas between floors as independent layers for floor detection, which is neglected in the previous studies. They can be detected by following certain movement patterns. During the movement from one floor to another in this experiment, the user needs to pass two staircases and a transition area between them. This needs to be treated as a whole process in floor detection. This paper first separately tests two methods: taking average or linearity. It recognizes that neither of these two methods can detect floor changes perfectly. Taking linearity is good at detecting sudden vertical movements but bad at dealing with the variations in the flat floors due to high sensitivity to slope changes. In the processing results in Fig.8b, it can be found that there is an extra detection on the

fourth floor when missing detections in third and second floors. On the other hand, taking average is more robust at detecting the horizontal movement but not able to detect vertical movements in time. As shown in Fig.8a, the detections of transition periods are always longer than that in reality. By integrating the results of these two methods, the transition area between floors requires to pass two changes of linearity and three different means of heights (Fig.8).

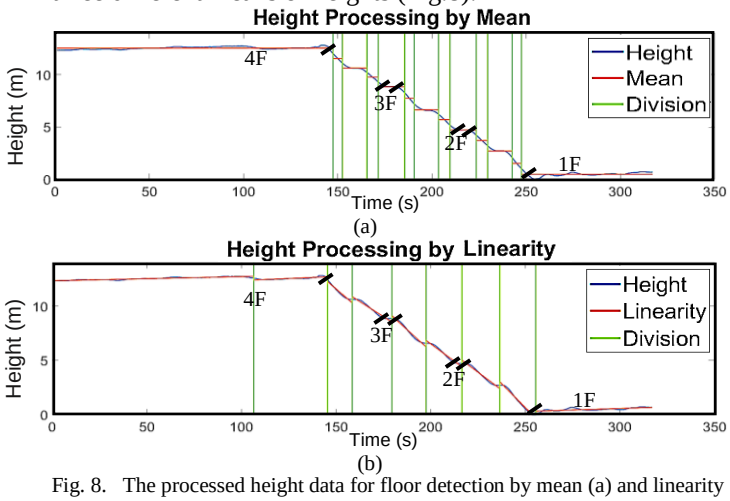


Fig. 8. The processed height data for floor detection by mean (a) and linearity (b) respectively (the real divisions between floors are marked out in black).

With this method, the floor detection will be related to real-time measurements by the number of detected floor changes and the initial floor level. These values can be determined by the average of in each height range of corresponded floor and they can be later used as a reference (Table I). The barometer data will be integrated with 2D PVINS data by limiting $(t(k) - t(i)) \rightarrow 0$, where $t(k)$ is the time steps of barometer readings.

TABLE I
HEIGHT RANGES FOR EACH FLOOR

Floor Number	Height Range (m)
4	>= 13.07
3.5[a]	9.03-13.07
3	8.85-9.03
2.5[a]	4.90-8.85
2	4.72-4.90
1.5[a]	0.70-4.72
1	<=0.70

[a] represents the transition area between every two floors.

IV. EXPERIMENTAL SETUP

The test site in this study is located at a four-floor building at University of Nottingham, Ningbo, China. All data are transferred to a desktop by wireless network for post-processing by MATLAB. The reference maps in WGS84 are the digitized floor plans imported to ArcGIS 10.3. They are posted on a web map repository using ArcGIS Online for indoor-outdoor transition, with the simple semantic representations of indoor structures. The overall length of the 2D reference path is approx. 161.57m and the total height of stairs is approx. 13.07m. The cameras are located on 4th floor in front of Room 416 and 1st floor in front of elevators, facing directly to the corresponding corridors with the targeted user in the center of the frame (Fig.9) and the height of cameras in the building are all in 3m.

For smartphone-based PDR system, a Huawei Mate8 (Android) in Android 6 and an iPhone7 Plus (iOS) in iOS 11 are used. The data collection app is MATLAB Mobile, which can achieve online data uploading during movement. The sampling frequency for two smartphones is set to be 50 Hz. During the experiment, both smartphones are held horizontally, pointing towards the walking directions. For the visual tracking system, the resolution of each camera is 680×540, the vertical FOV is 27°, and so the focal pixel length is about 1.05×103 per inch. The filming frequency is 17 frames per second (FPS). Cameras start filming simultaneously with the initialization of smartphone-based PDR. For floor detection, the barometer apps for pressure data collection are Barometer and Barograph. The latter is used for continuous recording and its sampling frequency is 1s-1. Barometers are triggered before the smartphone PDR for self-calibration and their timestamps will be recorded for later sensor fusion.

Fig. 9. 2D reference paths visualized from different floors with positions of cameras (a)-(c), 3D view of entire indoor path (d), and the webmap with outdoor environment in ArcGIS Online (e) (where 'adpt' represents the rooms other than offices and 'con' represents stairs and elevators).

V. RESULTS AND ANALYSIS

A. 2D Visual Tracking

In this study, the overall length of 2D visible paths is 56.13m,

which only accounts for 34.7% for the overall path. In previous studies, the non-occlusion path occupies at least 50% of the overall path when reaching decimeter-level accuracy [22-24, 51, 52], even some of them not reach completely invisible occasion but only partial occlusion [22-24]. This paper aims to validate whether the system can still work under this extreme condition.

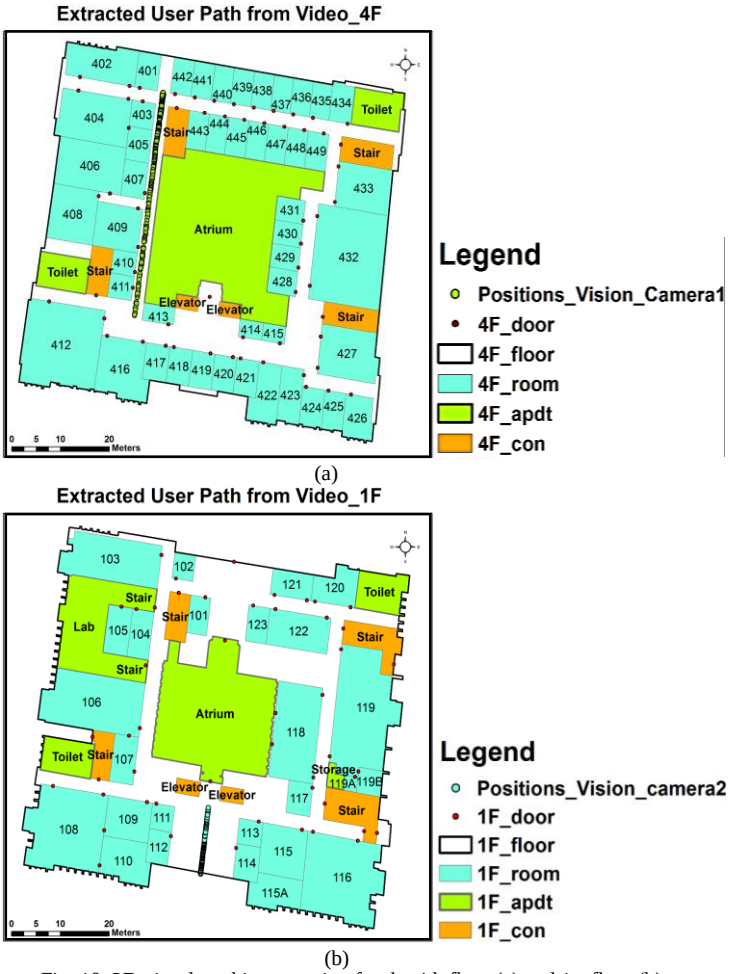


Fig. 10. 2D visual tracking mapping for the 4th floor (a) and 1st floor (b).

The OPS first provides the position of the functioning camera, which can also help for floor level calibration in LoS areas. As shown in Fig.10, with the mapping results from visual tracking, two partial paths from two cameras are matched well against the reference path. However, both two paths have a problem of unevenly distributed visual positioning points, though this phenomenon for the positioning points provided by the second camera is not distinct to be realized. This is mainly due to pinhole-based depth information calculation, which relies on h_i changes in frames. In the initial stage, the changes of h_i are trivial, leading to a dense distribution of step points; while in the ending part, the changes of h_i are becoming more significant [51, 52], leading to a more scattered distribution of positions. The length of the second partial path is quite short which will not raise great changes of h_i. Meanwhile, long distance between target and camera will also lead to mistakes in pedestrian detection.

Meanwhile, the filming frequency cannot match with the step frequency and the detected target positions are always in the middle of a step but cannot identify the starting and ending points of each step event. The previous visual gait detection [22-24] is not suitable for this study as: a) this paper does not apply foreground masks, which is quite labor intensive and responds slowly, while using a pinhole model for distance estimation; b) the filming frequency is lower than previous studies; and c) the ratio between IMU sampling frequency and filming frequency is not an integer. This makes the results from visual positioning more time-domain based rather than gait-based, and these data are not suitable to be directly used for calibrating the PDR positioning in visible areas, though it has an MAE of 0.06m. However, this will not affect the headings between steps and the information can be later applied for PDR calibration.

B. 2D Smartphone-Based PDR

1) 2D Calibration

In this experiment, the user walks 312 steps from the fourth floor to the first floor and PDR only provides the horizontal positions to avoid imposing additional errors due to the inclusion of the third dimension. The Android-running phone detects 296 steps while iOS-based phone detects 307 steps (Table II). The measurements have been repeated for 10 times, but the average detected steps do not change significantly, within 1 or 2 steps' fluctuations. This may be caused by data logging mechanism of PDR data, as it needs the network connection for data transmission while the RSS of Wireless Local Area Network (WLAN) is unstable in the experimental site. This could be resolved by using 4/5G for Mobile Communications as a supplement or using an offline system for data collection. The other reason to that may be due to a relatively higher sensitivity of the accelerometers embedded in iPhone7 Plus than that of Huawei Mate8's, providing a better step detection performance for the iPhone.

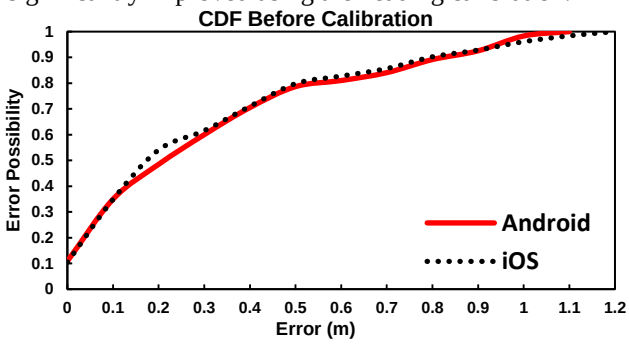
TABLE II
STEP DETECTION COMPARISON BETWEEN TWO TYPES OF SMARTPHONES

		Device	
	Actual Steps	iPhone7 Plus	Huawei Mate8
Detected Steps	312	307	296
Accuracy	100%	98.4%	94.9%

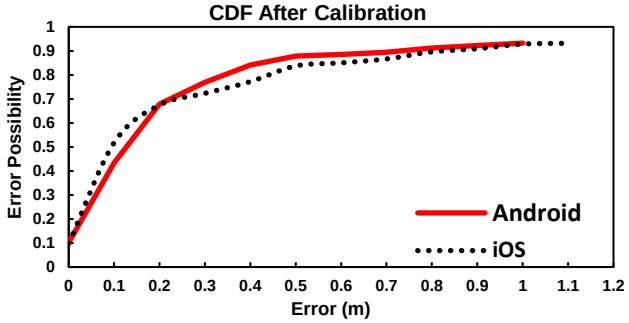
Before the heading calibration in the horizontal direction, the positional accuracies of these two types of smartphones are almost the same, i.e. the MAE is 0.31m (Android) and 0.29m (iOS) (Table III). However, according to their Cumulative Distribution Function (CDF) of error distributions, the maximum error of iOS is higher than that of Android's while it has more positioning points with an error less than 1m (Fig.11a). However, neither performs well enough for the staircase area with the frequent turnings (Fig.12a). This may be improved with the more sensitive gyroscope in the future with the advancement of embedded smartphone-based sensors.

After the 2D heading calibration, the MAEs of both types of smartphone-based PDR have been improved to 0.16m (Table III). 95% positioning points' error falls below 0.65m while before calibration it was 0.90m (Fig.11b). The Android-based-PDR seems to perform better after horizontal calibration without considerations of missing detected steps (Fig.12b). It

may be explained by the fact that more detected steps from the iOS system will introduce more difficulties to 2D calibration as this time when LoS areas only occupied 35.9% of the entire path. Therefore, the positioning accuracy cannot be significantly improved using the heading calibration.

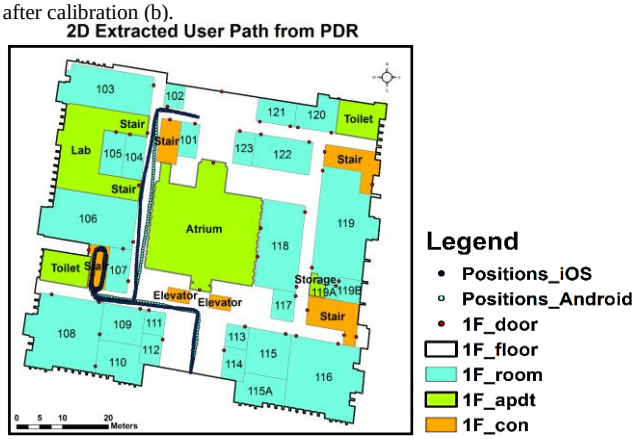


(a)

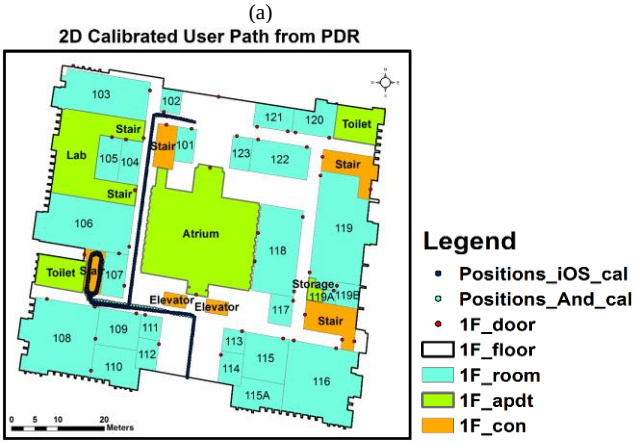


(b)

Fig. 11. The CDF distribution for 2D smartphone-based PDR before (a) and after calibration (b).



(a)



(b)

Fig. 12. The 2D projection on the 1st floor with all 2D PDR data before calibration (a) and after calibration (b).

TABLE III
POSITIONING ACCURACY ANALYSIS BEFORE AND AFTER CALIBRATION

Index for Accuracy		Device: Huawei Mate 8	Device: iPhone7 Plus
Pre-Calibration	MAE (m)*	0.31	0.29
	RMSE (m) *	0.43	0.43
Post-Calibration	MAE (m)	0.16	0.16
	Improvement	48.3%	44.8%
	RMSE (m) *	0.24	0.28
	Improvement	44.2%	34.8%

RMSE* is the Root Mean Squared Error, can be calculated as: $RMSE = \sqrt{MAE^2 + Variance}$.

2) Improvements to Other Studies Using 2D PVINS

Comparing with the recent study conducted by Shanghai Technology University [24], its MAE of PVINS can achieve 0.2m under the condition of partial occlusion, because it has more than 50% of positioning areas are in view during the experiment. However, the MAE proposed by this paper performs better, even with a simpler data fusion algorithm and a lower portion of non-occlusion areas. The CDFs in this paper also has an advantage of lower variations with about 90% of errors less than 0.4m, while this number for other studies is up to 0.5m. The other studies conducted by Missouri University [22, 23] have a minimum MAE of 0.5m, with a maximum improvement of 22.6%, summarizing from four tests. This is lower than the results achieved in this paper with a maximum improvement of 48.3% and an MAE at 0.16m (Table IV). Moreover, our designed referential path has 16 turnings during movement, which could easily introduce more errors due to the bias drift from the gyroscope. While in the other studies, the maximum number of turnings is 4. This shows its potential to deal with a path with a more complicated structure. Meanwhile, the ending points of both calibrated paths match well with the entrance of the building, which can keep tracking user trajectory and later be directly shifted to the outdoor positioning system with available GPS signals. This system also has been tested on two types of smartphones while the previous studies mainly focus on Android-running systems. However, the system proposed by this paper cannot provide gait-feature based visual tracking as in [22-24], which makes the visual tracking results not fully compatible with step event mechanism of PDR. It makes this system more single-directional calibration based, i.e. calibrating PDR by visual tracking but not the other way round.

TABLE IV
POSITIONING ACCURACY COMPARISON BETWEEN OTHER 2D PVINS STUDIES

Reference	[22, 23]	[24]	This Study
FPS	30	20	17
Visible Area (%)	>50	51.7	34.7
Device	Samsung Galaxy S4	Huawei Honor8	Huawei Mate8, iPhone7 Plus
Turnings	4	3	16
Best MAE (m)	0.50	0.20	0.16
Calibration	Bi-directional	Bi-directional	Single-directional

3) Comparison to Other Studies Using Magnetometer-Based Heading Calibration

This paper also compares its performance of 2D aspect to some other studies, who investigate alternative approaches, to

improve the performances of Dead Reckoning (DR) based Indoor Inertial Systems [15-21, 25, 26]. They use magnetometers for heading calibration instead of passive OPS. The majority of these studies apply self-developed hardware-suite without support from additional sensing system for precise positioning. Their preferred position for sensor wearing is on the foot, such as [15-20]. Their algorithms are mainly based on Zero-Velocity Updates (ZUPT) [15-20], with data fusion and error control based on EKF [16, 20] or Complimentary Filter (CF) [15, 19]. Some of the studies even integrate ZUPT and Step-and-Heading System (SHS) together for position estimation [17, 18]. Others hold the device in hand [25, 26] or put it on the waist [21]. These studies apply SHS for position estimations, with either Peak-Detection-based (PDT) or Zero-Detection-based (ZDT) algorithm for step detection. Among these methods, the accuracy and experimental conditions of these studies are shown in Table V. As almost all path types in other studies are close-loops and their accuracies are evaluated as Start-to-End radial distance, thus the positioning accuracy for this study is treated as the ratio between the largest error during estimation and total walking distance for reference. The distance error is also treated as the ratio between the overall length of the predicted distance and the ground truth due to the different lengths of the designed paths.

When comparing the positioning accuracy, the performance of the system in this case is not as good as these relatively precise foot-mounted INS systems with commercial IMU sensors, such as 0.3% in [20] and 0.4% in [15], though they are still comparable. This can be explained by the following reasons. First, the results are only calibrated in the LoS areas while the accuracy listed in Table V is the overall performance of both the visible and the invisible areas. The positioning accuracy in the LoS areas is 0.06%, which is much better than that in [15, 20]. The largest error actually appears in the invisible areas with frequent turnings, by only using smartphone-based PDR. This can be affected by the precision of applied hardware. The precision of the IMU sensors embedded in smartphones is not comparable to that of commercial foot-mounted sensors [4]. Moreover, as our system is tested on an open-loop path, it cannot have reverse-calibration as testing on a close-loop path [4]. Meanwhile, the foot-mounted systems have higher accuracy of step detection due to the position of sensor installation and the mechanism of ZUPT. However, our system has an advantage of higher accessibility as it only requires to have a specific smartphone app for data collection and transfer, while the foot-mounted systems with comparable accuracies [15, 20] require to wear specific body-attached sensors, cables, or batteries. Meanwhile, considering the user experience, it can be hard to persuade the users to wear specific sensor suites on body as in [15, 20] while our system only requiring current buildings to install a surveillance system. Although it also needs the potential costs of camera installation and calibration for the application, it may not be a problem as the installation of surveillance cameras is necessary not only for the tracking but also for the security purpose and the calibration is required only once. In addition, the surveillance system installation will be a ubiquitous requirement for the future buildings, which shows potential market for our system. Moreover, it shows a relatively higher accuracy of total distance estimation by using the camera calibration (0.1%) than all previous studies. For processing algorithm, the computation cost for deep learning is higher than that for EKF, however, this will be compensated by its high accuracy in the LoS areas (0.06%).

When compared to [26] with better overall performance among SHS-based systems, the system in this study has comparable performance on positioning accuracy. However, for the step detection, the accuracy of this system (98.4%) is slightly lower than that of [26] (98.67%). This may be also partially due to the hardware precision as mentioned above. Moreover, [26] divides the steps modes into four classes by SVM classification and introduces a Band-Pass Filter (BPF) for step detection under different walking modes. This may require more manual preparations before the test. However, our system does not have this process and just treats the whole process with one mixed class. Moreover, the difference between step-detection accuracies is not significant and the performance of our system is acceptable for positioning. Another advantage of our handheld system is the deployed sensor suite is already available in daily life and will be more easily accepted by user.

TABLE V
POSITIONING ACCURACY COMPARISON BETWEEN MAGNETOMETER-BASED STUDIES

Reference	[20]	[16]	[19]	[15]	[17]	[18]	[21]	[25]	[26]	This Study
Algorithm	ZUPT-EKF	ZUPT-EKF	ZUPT-CF	ZUPT-CF	ZUPT-SHS	ZUPT-SHS	SHS-PDT	SHS-ZDT	SHS-(BPF)PDT	SHS-ZDT
PDR Data Collection Devices	InertiaCube3	Self-created prototype	MicroStrain 3DM-GX1	MTi & MTi-G	Nano IMU	ADIS16405	NavMote	Nexus S	Self-created prototype	Huawei Mate8/ iPhone7 Plus
Device Positions	On Foot	On Foot	On Foot	On Foot	On Foot	On Foot	On Waist	Handheld	Handheld	Handheld
Path Type	Close loop	Close loop	Close loop	Close loop	Close loop	Close loop	Close loop	Close loop	Close loop	Open loop
Total Distance (m)	118.5	239.9	437.50	80	60	132	400	120	400	161.57
2D Positioning Accuracy (%)	0.3	2.01	1.0	0.4	2.0	3.26	3.0	4.2	1.8	0.6
2D Distance Error (%)	/	3.47	0.27	/	2.0	/	3.0	1.7-6.7	1.9	0.1
Data Transfer	Radio Frequency-Receiver	Bluetooth	Sony UXP-180 mini-computer	USB	Data Cable	Bluetooth	NetMote	USB	ARM Processor	WLAN

C. Height Estimation and Floor Detection

For 3D positioning, the common way to achieve that is to treat the horizontal and vertical localization separately [20, 30, 59, 66, 68-70, 88]. This may be due to the navigation mechanism, as the horizontal positioning is more important on each floor than in the transition areas in staircases and the vertical positioning only needs to provide the correct floor. However, as this study also considers the transition areas to be individual levels, it will both provide the height accuracy and floor detection accuracy for localization. Moreover, both the initial and the final floors have additional sensor information for floor level calibration, i.e. cameras' 3D locations in the main system. This can help improve the floor detection accuracy than using only the barometer-based floor detection algorithm.

1) Height Estimation and Floor Detection by Barometers

After the recorded pressure data are transferred into height, the MAEs of estimated height information from both types of smartphones is about 0.5m, which is not as good as that of using two barometers with one as a reference device with an accuracy of 0.15m [71]. However, it is better than the methods with a single barometer, which only achieves an accuracy of 1 to 2m [68-70, 88] (Table VI). Considering the low-cost and easy implementation, our method is still a better choice than other methods with comparable accuracy.

TABLE VI
ACCURACY COMPARISON BETWEEN OTHER STUDIES USING BAROMETERS

Reference	[88]	[68]	[69, 70]	[71]	This Study
Methods	BPF	Relative height fingerprint	Relative height fingerprint + GNSS signals	Reference Device	Self-calibration + Mean and Slope Change Detection
No. of Barometers	1	1	1	2	1
Device	Self-created prototype	Samsung Galaxy S3	Unknown Android Phone	Samsung Galaxy N5 Samsung Galaxy S4	Huawei Mate8/ iPhone7 Plus
MAE (m)	1.20	1~2	1~2	0.15	0.5

After being processed using the floor detection algorithm, the results show that the barometers from both types of smartphones are sensitive enough to recognize the floors with relatively high accuracy, i.e. 98%. The errors typically appear in the first stages of the movement going down from the stairs. This may be due to the imprecision of embedded barometers, as the previous studies have faced the similar problems as the minimum height change that can be detected is about 1.6m [57, 66, 89]. However, the height difference between every two stairs (0.16m) is smaller than that range in this experiment.

2) Comparison with IMU-based Height Estimation

Some studies explore the accuracy of using vertical acceleration changes based on foot-mounted INS for height estimation [16, 20]. As the experimental conditions of these studies are different, the accuracy will be assessed by the ratio between estimated height error and the overall height of the staircases (Table VII). The results suggest that the barometer-assisted height detection is comparable to these foot-mounted sensor systems, even using lower precision of embedded hardware in smartphones.

TABLE VII
ACCURACY COMPARISON BETWEEN OTHER STUDIES USING ACCELEROMETERS

Reference	[20]	[16]	This Study
Methods	ZUPT	ZUPT + Probabilistic Neutral Network Classification	Self-calibration + Mean and Slope Change Detection
Total Height (m)	3	7.84	13.07
Device	InertiaCube3	Self-created Prototype	Huawei Mate8/ iPhone7 Plus
Mean Accuracy (%)	2	6.42	3.8

D. 3D Localization and Comparison to Other Studies

A 3D path is produced after the integration with previous calibrated results of 2D PVINS by similar time stamps (Fig.13). However, as not all the steps are detected, there are some additional errors being introduced into PDR-based positioning system besides the errors from the barometer measurements, especially for Android-running system as it has more undetected steps. Moreover, as the step event frequency is not perfectly matching with that of height data, which will be another error source for the 3D localization. Thus, 3D positions estimated by Android-running system will have larger total MAE (1.55m) than that by the iOS-based system (1.52m). The errors mainly come from the transition areas, where there is no calibration from visual positioning and the barometer cannot deal with the quick changes of insignificant changes of height by walking downstairs (Fig.13), which has also been proved by [61] with similar conclusions.

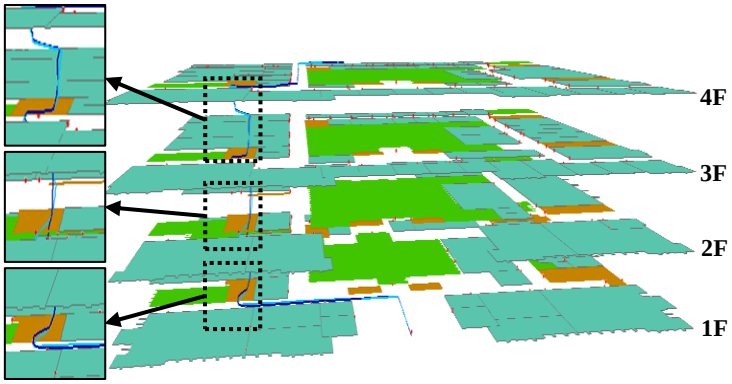


Fig. 13. The 3D view of the estimated path by smartphone-based PDR (Android in light blue and iOS in dark blue) and the locations of main errors in dash boxes.

When comparing to the other systems with precise IMU sensors [20, 26], their performances are not affected by the missing detection of steps during sensor fusion. Therefore, their previous higher accuracies in both 2D positioning and height estimation will lead to a relatively better 3D positioning accuracy, with 0.3% in [20], and 1.1% in [26] (the accuracy here is the ratio between estimated error and the total distance of referential path). The accuracy of the proposed system is about 0.9%, which can be regarded as comparable to these studies. Moreover, this is also better than the previous study using multi-sensor system including Wi-Fi, iBeacons, and barometer for positioning with a 3D positioning accuracy of

1.7% [61], while having no additional cost for installation or infrastructure management. The accuracy of the proposed system may be improved in future with the PDR algorithm or the advancement of embedded IMU sensors to have higher sensitivity to detect the correct number of steps.

However, the requirement of 3D positioning accuracy is less important for real applications as it usually requires 2.5D positioning instead of real 3D positioning. The user positions can then be represented as $P^*(x_R, y_R, J)$ by providing the horizontal positions (x_R, y_R) and the correct floor number J. By integrating floor number information into the previous 2D system based similar time stamps, the overall performance of the system will not be significantly affected as this time the 2D positions are more important and the floor detection accuracy is high enough to handle automatic floor plan changes.

E. Limitations before Developing into Real-Time System

Like other similar studies [15-26], the data in this study is post-processed after transmission to the desktop. This is mainly limited by the visual data acquisition due to the privacy policy in the university and the visual data is not allowed to be transmitted to the desktop in real time. Meanwhile, the 'PDR' and the 'Height Estimation' sub-systems have already achieved real-time processing as the inertial and pressure data can be sent to desktop and processed during the movement via WLAN and the positions of the user will be stored in the system. The current offline system can be used for low-cost 3D mobile mapping, which can help calibrate the moving trajectories for 2D laser scanning to build 3D indoor models. It also can provide historical paths of the indoor pedestrians for security checking. In the future, one of the limitations of turning this system into an online system will be the live streaming speed of surveillance videos. This is determined by the available bandwidth of the existing WLAN in the building. For the current system, the bandwidth should be approx. 6 Mbps for each camera, while the university's WLAN bandwidth is 10 Mbps and it can fully support its live streaming. The storage of the data may be another problem. However, this system is designed for a whole building with a powerful processing center and it is assumed to finish all processing in the mainstream and send the data back to the user's device via the network, like the idea mentioned in [21]. The requirement of the computation power for real-time detection is not very high. In this study, the computer has a CPU in Intel Core i7-7700, a GPU in NVIDIA GTX 1080, and 16G RAM, which is commonly used in the field of computer vision industry.

VI. CONCLUSION

This study has designed a novel low-cost and user-friendly 3D PVINS that uses multi-cameras, smartphone-based PDR and embedded barometer, and provides a comparable 3D accuracy of 0.9%. The novelty of this system is: (a) a modified Faster R-CNN based passive visual tracking, with simple implementation, high accuracy, and real-time detection; (b) a novel algorithm for multi-scene shifting with automatic PDR turning detection; (c) a novel data fusion method with simple operation and high effectiveness, achieving more than 20% 2D accuracy improvement for severe occlusion-affected areas than the previous 2D PVINSs; d) a novel algorithm for height/floor estimation with more detailed floor-level division using single embedded barometer in a smartphone; e) the acquired results with absolute coordinates to be directly used in outdoor systems; f) the application on both Android-running and iOS-running smartphones with better robustness than previous Android-only systems. This system can provide 2D positions of each floor with an accuracy of 0.16m while identifying the current floor level of the users with 98% detection accuracy (0.5m vertical accuracy), which has already reached the requirement by FCC, with 50m horizontal accuracy and 3m vertical accuracy [54]. Another advantage of this 3D PVINS is no special requirement of attaching instruments on user bodies or using specific sensor-suite as settlements in other self-contained systems, which makes them more accessible and user-friendly for the future applications. However, the PDR algorithm used in this study needs further improvement, because there are more missing steps with the accumulation of distance. This may be due to the data logging mechanism, and it may be solved by temporary data storage on a user's device and resuming data transmission when having Wi-Fi connection again. Moreover, as this system is currently designed for single user tracking, it still has the potential to be developed into a multi-user system, which needs to improve the algorithm of visual tracking. The acquisition of surveillance data may be another limitation before turning the current system into a real-time system as it will raise the issue of personal privacy and this time the permission is pre-applied for data downloading. The floor identification approach can also be more precise to identify exact user 3D locations inside buildings.

REFERENCES

[1] W. Elloumi, A. Latoui, R. Canals, A. Chetouani, and S. Treuillet, "Indoor pedestrian localization with a smartphone: a comparison of inertial and vision-based methods," IEEE Sensors Journal, vol. 16, no. 13, pp. 5376-5388, 2016.

[2] T. C. Dong-Si and A. I. Mourikis, "Estimator initialization in vision-aided inertial navigation with unknown camera-IMU calibration," in IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1064-1071, 2012.

[3] C. Fuchs, N. Aschenbruck, P. Martini, and M. Wieneke, "Indoor tracking for mission critical scenarios: A survey," Pervasive and Mobile Computing, vol. 7, no. 1, pp. 1-15, 2011.

[4] R. Harle, "A survey of indoor inertial positioning systems for pedestrians," IEEE Communications Surveys and Tutorials, vol. 15, no. 3, pp. 1281-1293, 2013.

[5] J. Racko, P. Brida, A. Perttula, J. Parviainen, and J. Collin, "Pedestrian dead reckoning with particle filter for handheld smartphone," in 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-7, 2016.

[6] A. Basiri, E. S. Lohan, T. Moore, A. Winstanley, P. Peltola, C. Hill, et al., "Indoor location based services challenges, requirements and usability of current solutions," Computer Science Review, 2017.

[7] J. P. Tardif, M. George, M. Laverne, and A. Kelly, "A new approach to vision-aided inertial navigation," in IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4161-4168, 2010.

[8] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in 2007 IEEE International Conference on Robotics and Automation, Roma, Italy, pp. 3565-3572, 2007.

[9] O. J. Woodman, "An introduction to inertial navigation," PhD Thesis, Computer Laboratory, University of Cambridge, 2007.

[10] S. Rajagopal, "Personal dead reckoning system with shoe mounted inertial sensors," Master's Degree Project, Stockholm, Sweden, 2008.

[11] K. Abdulrahim, C. Hide, T. Moore, and C. Hill, "Aiding low cost inertial navigation with building heading for pedestrian navigation," Journal of Navigation, vol. 64, no. 2, pp. 219-233, 2011.

[12] P. C. Lin, J. C. Lu, C. H. Tsai, and C. W. Ho, "Design and implementation of a nine-axis inertial measurement unit," IEEE/ASME Trans. Mechatronics, vol. 17, no. 4, pp. 657-668, 2012.

[13] D. Griesbach, D. Baumbach, and S. Zuev, "Stereo-vision-aided inertial navigation for unknown indoor and outdoor environments," International Journal of Cultural Policy, vol. 20, no. 3, pp. 281-295, 2014.

[14] J. A. B. Link, P. Smith, N. Viol, and K. Wehrle, "FootPath: Accurate map-based indoor navigation using smartphones," in 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-8, 2011.

[15] H. Fourati, "Heterogeneous data fusion algorithm for pedestrian navigation via foot-mounted inertial measurement unit and complementary filter," IEEE Trans. Instrumentation and Measurement, vol. 64, no. 1, pp. 221-229, 2015.

[16] Y.-L. Hsu, J.-S. Wang, and C.-W. Chang, "A wearable inertial pedestrian navigation system with quaternion-based extended Kalman filter for pedestrian localization," IEEE Sensors Journal, vol. 17, no. 10, pp. 3193-3206, 2017.

[17] C. Huang, Z. Liao, and L. Zhao, "Synergism of INS and PDR in self-contained pedestrian tracking with a miniature sensor module," IEEE Sensors Journal, vol. 10, no. 8, pp. 1349-1359, 2010.

[18] X. Meng, Z.-Q. Zhang, J.-K. Wu, W.-C. Wong, and H. Yu, "Self-contained pedestrian tracking during normal walking using an inertial/magnetic sensor module," IEEE Trans. Biomedical Engineering, vol. 61, no. 3, pp. 892-899, 2014.

[19] X. Yun, J. Calusdian, E. R. Bachmann, and R. B. McGhee, "Estimation of human foot motion during normal walking using inertial and magnetic sensor measurements," IEEE Trans. on Instrumentation and Measurement, vol. 61, no. 7, pp. 2059-2072, 2012.

[20] E. Foxlin, "Pedestrian tracking with shoe-mounted inertial sensors," IEEE Computer Graphics and Applications, vol. 25, no. 6, pp. 38-46, 2005.

[21] L. Fang, P. J. Antsaklis, L. A. Montestruque, M. B. McMickell, M. Lemmon, Y. Sun, et al., "Design of a wireless assisted pedestrian dead reckoning system-the NavMote experience," IEEE Trans. Instrumentation and Measurement, vol. 54, no. 6, pp. 2342-2358, 2005.

[22] W. Jiang and Z. Yin, "Combining passive visual cameras and active IMU sensors to track cooperative people," in 18th International Conference on Information Fusion (FUSION), pp. 1338-1345, 2015.

[23] W. Jiang and Z. Yin, "Combining passive visual cameras and active IMU sensors for persistent pedestrian tracking," Journal of Visual Communication and Image Representation, vol. 48, pp. 419-431, 2017.

[24] J. Zhang and P. Zhou, "Integrating low-resolution surveillance camera and smartphone inertial sensors for indoor positioning," in 2018 IEEE/ION Position, Location and Navigation Symposium (PLANS), pp. 410-416, 2018.

[25] N. Kothari, B. Kannan, E. D. Glasgwow, and M. B. Dias, "Robust indoor localization on a commercial smart phone," Procedia Computer Science, vol. 10, pp. 1114-1120, 2012.

[26] H. Zhang, W. Yuan, Q. Shen, T. Li, and H. Chang, "A handheld inertial pedestrian navigation system with accurate step modes and device poses recognition," IEEE Sensors Journal, vol. 15, no. 3, pp. 1421-1429, 2015.

[27] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, "Recommendations in location-based social networks: A survey," GeoInformatica, vol. 19, no. 3, pp. 525-565, 2015.

[28] F. Bentley, H. Cramer, and J. Müller, "Beyond the bar: The places where location-based services are used in the city," Personal and Ubiquitous Computing, vol. 19, no. 1, pp. 217-223, 2015.

[29] M. Zhang, J. D. Hol, L. Slot, and H. Luinge, "Second order nonlinear uncertainty modeling in strapdown integration using MEMS IMUs," in 2011 Proceedings of the 14th International Conference on Information Fusion (FUSION), pp. 1-7, 2011.

[30] A. M. Sabatini and V. Genovese, "A sensor fusion method for tracking vertical velocity and height based on inertial and barometric altimeter measurements," Sensors, vol. 14, no. 8, pp. 13324-13347, 2014.

[31] G. Panahandeh and M. Jansson, "Vision-aided inertial navigation based on ground plane feature detection," IEEE/ASME Trans. Mechatronics, vol. 19, no. 4, pp. 1206-1215, 2014.

[32] J. Pinchin, C. Hide, and T. Moore, "The use of high sensitivity gps for initialisation of a foot mounted inertial navigation system," in 2012 IEEE/ION Position Location and Navigation Symposium (PLANS), pp. 998-1007, 2012.

[33] A. Vu, A. Ramanandan, A. Chen, J. A. Farrell, and M. Barth, "Real-time computer vision/DGPS-aided inertial navigation system for lane-level vehicle navigation," IEEE Trans. Intelligent Transportation Systems, vol. 13, no. 2, pp. 899-913, 2012.

[34] S. Adler, S. Schmitt, K. Wolter, and M. Kyas, "A survey of experimental evaluation in indoor localization research," in 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-10, 2015.

[35] R. Mautz and S. Tilch, "Survey of optical indoor positioning systems," in 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-7, 2011.

[36] C. He, P. Kazanzides, H. T. Sen, S. Kim, and Y. Liu, "An inertial and optical sensor fusion approach for six degree-of-freedom pose estimation," Sensors, vol. 15, no. 7, pp. 16448-16465, 2015.

[37] B. Hartmann, N. Link, and G. F. Trommer, "Indoor 3D position estimation using low-cost inertial sensors and marker-based video-tracking," in 2010 IEEE/ION Position Location and Navigation Symposium (PLANS), pp. 319-326, 2010.

[38] A. Roy, P. Chattopadhyay, S. Sural, J. Mukherjee, and G. Rigoll, "Modelling, synthesis and characterisation of occlusion in videos," IET Computer Vision, vol. 9, no. 6, pp. 821-830, 2015.

[39] B. H. Yuan, D. X. Zhang, K. Fu, and L. J. Zhang, "Video tracking of human with occlusion based on Meanshift and Kalman filter," in Applied Mechanics and Materials, pp. 3672-3677, 2013.

[40] M. Mirabi and S. Javadi, "People tracking in outdoor environment using Kalman filter," in 2012 Third International Conference on Intelligent Systems Modelling and Simulation, pp. 303-307, 2012.

[41] B. De Villiers, W. Clarke, and P. Robinson, "Mean shift object tracking with occlusion handling," in 21st Conference of the International Association for Pattern Recognition (IAPR), 2012.

[42] J. Yan, Q. Ling, Y. Zhang, F. Li, and F. Zhao, "A novel occlusion-adaptive multi-object tracking method for road surveillance applications," in 2013 32nd Chinese Control Conference (CCC), pp. 3547-3551, 2013.

[43] Y. Hua, K. Alahari, and C. Schmid, "Occlusion and motion reasoning for long-term tracking," in European Conference on Computer Vision, pp. 172-187, 2014.

[44] G. C. Barceló, G. Panahandeh, and M. Jansson, "Image-based floor segmentation in visual inertial navigation," in 2013 IEEE International Instrumentation and Measurement Technology Conference, Minneapolis, MN, United States, pp. 1402-1407, 2013.

[45] G. Panahandeh, D. Zachariah, and M. Jansson, "Exploiting ground plane constraints for visual-inertial navigation," in 2012 IEEE/ION Position Location and Navigation Symposium (PLANS), pp. 527-534, 2012.

[46] A. Ramanandan, A. Chen, and J. A. Farrell, "Inertial navigation aiding by stationary updates," IEEE Trans. Intelligent Transportation Systems, vol. 13, no. 1, pp. 235-248, 2012.

[47] X. Song, L. D. Seneviratne, and K. Althoefer, "A Kalman filter-integrated optical flow method for velocity sensing of mobile robots," IEEE/ASME Trans. Mechatronics, vol. 16, no. 3, pp. 551-563, 2011.

[48] C. Hide, T. Botterill, and M. Andreotti, "Vision-aided IMU for handheld pedestrian navigation," in GNSS 2010 Conference Proceedings of the Institute of Navigation, Portland, Oregon, 2010.

[49] C. Hide, T. Botterill, and M. Andreotti, "Low cost vision-aided IMU for pedestrian navigation," in Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), pp. 1-7, 2010.

[50] M. Li, B. H. Kim, and A. I. Mourikis, "Real-time motion tracking on a cellphone using inertial sensing and a rolling-shutter camera," in 2013 IEEE International Conference on Robotics and Automation (ICRA), pp. 4712-4719, 2013.

[51] J. Yan, G. He, A. Basiri, and C. Hancock, "Vision-aided indoor pedestrian dead reckoning," in 2018 IEEE International Instrumentation & Measurement Technology Conference, Houston, USA, pp. 1-6, 2018.

[52] J. Yan, G. He, A. Basiri, and C. Hancock, "Indoor pedestrian dead reckoning calibration by visual tracking and map information," in Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS), pp. 1-10, 2018.

[53] J. Yan, G. He, and C. Hancock, "Low-cost vision-based positioning system," in Adjunct Proceedings of the 14th International Conference on Location Based Services, pp. 44-49, 2018.

[54] FCC, "Fourth Report and Order in PS Docket No. 07-114," Federal Communications Commission (FCC)2015.

[55] M. Tanigawa, H. Luinge, L. Schipper, and P. Slycke, "Drift-free dynamic height sensor using MEMS IMU aided by MEMS pressure sensor," in 5th Workshop on Positioning, Navigation and Communication, pp. 191-196, 2008.

[56] X. Shen, Y. Chen, J. Zhang, L. Wang, G. Dai, and T. He, "BarFi: Barometer-aided Wi-Fi floor localization using crowdsourcing," in 2015 IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems, pp. 416-424, 2015.

[57] H. Ye, T. Gu, X. Tao, and J. Lu, "Scalable floor localization using barometer on smartphone," Wireless Communications and Mobile Computing, vol. 16, no. 16, pp. 2557-2571, 2016.

[58] M. Zhang, A. Vydhyanathan, A. Young, and H. Luinge, "Robust height tracking by proper accounting of nonlinearities in an integrated UWB/MEMS-based-IMU/baro system," in 2012 IEEE/ION Position Location and Navigation Symposium (PLANS), pp. 414-421, 2012.

[59] H. Ye, T. Gu, X. Zhu, J. Xu, X. Tao, J. Lu, et al., "FTrack: Infrastructure-free floor localization via mobile phone sensing," in 2012 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 2-10, 2012.

[60] I. Constandache, X. Bao, M. Azizyan, and R. R. Choudhury, "Did you see bob?: Human localization using mobile phones," in Proceedings of the 16th Annual International Conference on Mobile Computing and Networking, pp. 149-160, 2010.

[61] F. Ebner, T. Fetzer, F. Deinzer, L. Köping, and M. Grzegorzek, "Multi-sensor 3D indoor localisation," in 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-11, 2015.

[62] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury, "No need to war-drive: Unsupervised indoor localization," in Proceedings of the 10th InternationalCconference on Mobile Systems, Applications, and Services, pp. 197-210, 2012.

[63] B. Li, B. Harvey, and T. Gallagher, "Using barometers to determine the height for indoor positioning," in 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-7, 2013.

[64] H. Xia, X. Wang, Y. Qiao, J. Jian, and Y. Chang, "Using multiple barometers to detect the floor location of smart phones with built-in barometric sensors for indoor positioning," Sensors, vol. 15, no. 4, pp. 7857-7877, 2015.

[65] H. Wang, H. Lenz, A. Szabo, U. D. Hanebeck, and J. Bamberger, "Fusion of barometric sensors, wlan signals and building information for 3D indoor/campus localization," in Proceedings of International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), pp. 426-432, 2006.

[66] K. Muralidharan, A. J. Khan, A. Misra, R. K. Balan, and S. Agarwal, "Barometric phone sensors: More hype than hope!," in Proceedings of the 15th Workshop on Mobile Computing Systems and Applications, pp. 1-6, 2014.

[67] J.-s. Jeon, Y. Kong, Y. Nam, and K. Yim, "An indoor positioning system using bluetooth RSSI with an accelerometer and a barometer on a smartphone," in 2015 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), pp. 528-531, 2015.

[68] J. Z. Flores and R. Farcy, "Indoor navigation system for the visually impaired using one inertial measurement unit (IMU) and barometer to guide in the subway stations and commercial centers," in International Conference on Computers for Handicapped Persons, pp. 411-418, 2014.

[69] T. Lin, L. Li, and G. Lachapelle, "Multiple sensors integration for pedestrian indoor navigation," in 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-9, 2015.

[70] B. Shin, S. Lee, C. Kim, J. Kim, T. Lee, C. Kee, et al., "Implementation and performance analysis of smartphone-based 3D PDR system with hybrid motion and heading classifier," in 2014 IEEE/ION Position, Location and Navigation Symposium-PLANS, pp. 201-204, 2014.

[71] S.-S. Kim, J.-W. Kim, and D.-S. Han, "Floor detection using a barometer sensor in a smartphone," in 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2017.

[72] W. Kang, S. Nam, Y. Han, and S. Lee, "Improved heading estimation for smartphone-based indoor positioning systems," in 2012 IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), pp. 2449-2453, 2012.

[73] P. Goyal, V. J. Ribeiro, H. Saran, and A. Kumar, "Strap-down pedestrian dead-reckoning system," in 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-7, 2011.

[74] H. Weinberg, "Using the ADXL202 in pedometer and personal navigation applications," Analog Devices AN-602 Application Note, vol. 2, no. 2, pp. 1-6, 2002.

[75] F. Danion, E. Varraine, M. Bonnard, and J. Pailhous, "Stride variability in human gait: the effect of stride frequency and stride length," Gait and Posture, vol. 18, no. 1, pp. 69-77, 2003.

[76] S. J. Mason, G. E. Legge, and C. S. Kallie, "Variability in the length and frequency of steps of sighted and visually impaired walkers," Journal of Visual Impairment and Blindness, vol. 99, no. 12, pp. 741-754, 2005.

[77] Y. Huang, O. G. Meijer, J. Lin, S. M. Bruijn, W. Wu, X. Lin, et al., "The effects of stride length and stride frequency on trunk coordination in human walking," Gait and Posture, vol. 31, no. 4, pp. 444-449, 2010.

[78] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," Computer Vision and Image Understanding, vol. 104, no. 2, pp. 90-126, 2006.

[79] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," IEEE Trans.Pattern Analysis and Machine Intelligence, no. 12, pp. 2179-2195, 2008.

[80] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 34, no. 4, pp. 743-761, 2012.

[81] T.-H. Tsai, C.-H. Chang, and S.-W. Chen, "Vision based indoor positioning for intelligent buildings," in 2016 2nd International Conference on Intelligent Green Building and Smart Grid (IGBSG), pp. 1-4, 2016.

[82] Y. Zhou, S. Zlatanova, Z. Wang, Y. Zhang, and L. Liu, "Moving human path tracking based on video surveillance in 3D indoor scenarios," ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 3, no. 4, pp. 97-101, 2016.

[83] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587, 2014.

[84] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in European Conference on Computer Vision, pp. 346-361, 2014.

[85] R. Girshick, "Fast R-CNN," in Proceedings of the IEEE International Conference on Computer Vision, pp. 1440-1448, 2015.

[86] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Advances in Neural Information Processing Systems, pp. 91-99, 2015.

[87] Y. Cai and X. Tan, "Weakly supervised human body detection under arbitrary poses," in 2016 IEEE International Conference on Image Processing (ICIP), pp. 599-603, 2016.

[88] K. Sagawa, H. Inooka, and Y. Satoh, "Non-restricted measurement of walking distance," in 2000 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1847-1852, 2000.

[89] H. Ye, T. Gu, X. Tao, and J. Lu, "B-Loc: Scalable floor localization using barometer on smartphone," in 2014 IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), pp. 127-135, 2014.

Jingjing Yan (M'18) received the B.S. degree in geographical science from University of Nottingham, Nottingham, UK, in 2013 and the M.S. degree in geography from University College London, London, UK, in 2014. She is currently pursuing the Ph.D. degree in geographical science at University of Nottingham Ningbo China.

Her current research interest includes low-cost indoor pedestrian positioning using techniques of PDR, visual tracking, floor detection, and action recognition.

Gengen He received the B.S. and M.S. degrees in biology from Georgetown University, Washington D.C., USA, in 2006 and 2007 and the Ph.D. degree in geography from University of Tennessee, Knoxville, USA in 2017.

Since 2016, he has been an Assistant Professor with the Geographical Sciences Department, University of Nottingham Ningbo China. His current research interests include development and application of geospatial technology to improve human navigation and environmental understanding, data creation in the indoor environment, and integration of optimized modeling algorithms for smartphones, drones, and robotics.

Anahid Basiri received the B.S. degree in civil engineering and geomatics in 2006, M.S. and Ph.D. degree in geospatial information science in 2008 and 2012, from K.N.Toosi University, Tehran, Iran.

From 2012 to 2013, she started a postdoc position in Maynooth University, Ireland on pedestrian navigation. From 2013 to 2016, she worked as a Marie Curie Fellow at the Nottingham Geospatial Institute, University of Nottingham. Since 2017, she is a Lecturer in Spatial Data Science and Visualization at the Centre for Advanced Spatial Analysis, University College London, UK. Her current research interest includes spatio-temporal data mining techniques for crowd-sourced geospatial data to understand patterns and behaviors of LBS users while the data may suffer from several aspects of uncertainty and bias.

Craig Hancock received the B.S. degree in surveying and mapping science and the Ph.D. degree in space geodesy respectively from Newcastle University, Newcastle, UK, in 2003 and 2012.

Currently, he is the Head of Civil Engineering at University of Nottingham, Ningbo campus. His current research interests include structural deformation monitoring using GNSS and remote sensing, mapping utilities in urban environments using GNSS and other location technologies, GNSS error mitigation with particular emphasis on ionospheric scintillation, and terrestrial laser scanning. He is a member of the editorial board for the journal "Survey Review".